Volume 240, number 1,2, 41–44 FEB 06481 November 1988

Occurrence and spacing of ribosome recognition sites in mRNAs of chloroplasts from higher plants

Michael Ruf and Hans Kössel

Institut für Biologie III der Universität Freiburg, Schänzle-Str. 1, D-7800 Freiburg, FRG

Received 9 September 1988

A computer-aided search for potential ribosome recognition sequences of mRNAs from tobacco chloroplasts shows that more than 90% of mRNA species contain sequences upstream of the respective initiator codons, which allow base pairing with 3'-terminal sequences of small subunit rRNA. This complementarity in several cases involves 16 S rRNA sequences between the canonical CCUCC sequence and the 3'-terminal stem/loop structure. The distances between potential ribosome recognition sequences and initiator codons can be up to 25 nucleotides which is much greater when compared to the spacing of 7 ± 2 nucleotides observed for the classical Shine-Dalgarno sequences in bacterial mRNAs.

Chloroplast; Translational initiation; Ribosome recognition site

1. INTRODUCTION

Small subunit rRNAs from chloroplasts are homologous to bacterial 16 S rRNA [1,2] and the strong conservation of their 3'-terminal sequences, in particular, has led to the proposal [3] that binding of plastidic mRNAs to small ribosomal subunits is mediated by a similar base pairing between the 3'-terminal sequences of small subunit rRNA and translational start sites of mRNA as is evident in bacterial systems [4–7]. A compilation of translational start sequences [8] shows that bacterial mRNAs contain a maximum of 9 consecutive bases complementary to the 3'-terminal sequences of 16 S rRNA and that they are positioned at a distance of 7 ± 2 nucleotides upstream of the respective initiator codons. Chloroplasts mRNAs, however, appear not to follow this rule consistently. Thus, while mRNAs coding for the large subunit of ribulosebisphosphate carboxylase from maize [9] and spinach [10] contain canonical Shine-Dalgarno-sequences, no such sequences could be identified in psbA mRNAs from tobacco [11] and spinach [12]. Our recent observation of a potential ribosome binding site in the rpoA mRNA of maize, which is positioned 18–22 nucleotides upstream of the AUG codon [13], has prompted us to carry out a computer-aided search for potential ribosome-binding sites of mRNAs encoded in the tobacco plastome. The compilation presented here indicates that the vast majority of mRNAs from tobacco chloroplasts do, indeed, contain sequences suitable for ribosome binding. In many cases, however, a spacing between these sequences and the respective initiator codons of up to 25 nucleotides is necessary to permit the mRNA-16 S rRNA base pairing interaction.

Correspondence address: M. Ruf, Institut für Biologie III der Universität Freiburg, Schänzle-Str. 1, D-7800 Freiburg, FRG

2. METHODS

For the computer-aided screening of sequences complementary to the 3'-terminal region of chloroplast 16 S rRNA, sequences upstream of the respective initiator codons as far as the positions −50 of the 39 identified protein genes from the tobacco plastome [14] were used. Unidentified open reading frames and the putative *ndh* genes were not included. The initiator codon of the *perD* gene was used in accordance with the short exon identified more recently [15]. The *rpoC* gene was included in its disected form ($rpoC_1$ and $rpoC_2$) according to the analysis of the two cistrons in spinach [16] and maize (unpublished, this laboratory).

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50 © 1988 Federation of European Biochemical Societies

Sequences with a minimum of 4 consecutive positions complementary to the 3′-terminal 14 nucleotides of small subunit rRNA were screened. GU pairs and base pairs involving the first G position of the 3′-terminal stem of 16 S rRNA (stem 48 according to [2], see top of fig.1) were included for the screening. In the case of several potential candidates, only the recognition sequence closest to the respective initiator codon was taken into account.

3. RESULTS AND DISCUSSION

In fig.1 the 3′-terminal sequence of chloroplast 16 S rRNA is depicted compared to an alignment of the leader sequences of rpoA mRNA from maize [13], spinach [17] and tobacco [14]. Potential recognition sequences of those leader regions (underlined in fig.1) comprise 5 consecutive nucleotides with a distance to the initiator codons of 17, 19 and 18 nucleotides, respectively. It should be noted that this complementarity does not include the canonical CCUCC sequence [8] of the 16 S rRNA. Instead of this, the neighbouring sequence, ACUAG, which is flanked by the 3′-terminal stem/loop structure (see fig.1) can be used to form appropriate base pairs. As a consequence, the canonical Shine-Dalgarno sequence, GGAGG on the mRNA, is not obligatory for complementarity (and hence for recognition) in this case. A similar situation was found for several of the leader regions listed in fig.2.

The result of our screening for potential ribosome recognition sites of mRNAs derived from the known sequence of the tobacco plastome

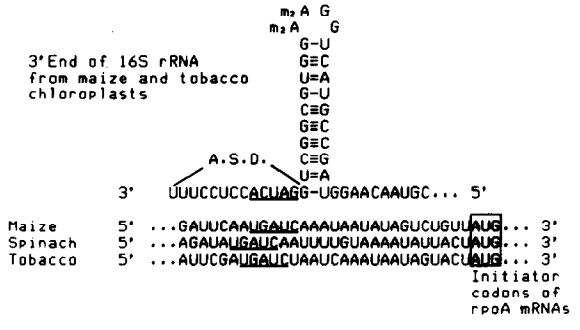

Fig.1. Complementarity of the 3′-terminal sequence from plastidic 16 S rRNA [3] with leader sequences of rpoA mRNAs from maize [13], spinach [17] and tobacco [14] chloroplasts. Potential recognition sequences are marked by underlining. For screening of complementarity with other mRNAs (see fig.2) the entire single-stranded end of 16 S rRNA as marked by A.S.D. (anti-Shine-Dalgarno sequence) was included.

is summarized in fig.2. Of the 41 protein genes identified [14–16], 37 genes listed in fig.2 show regions of complementarity with 3′-terminal sequences of 16 S rRNA. As indicated in the figure, the length of the potential recognition sites ranges from 4 (chosen as lower limit for the screening) to a maximum of 9 nucleotides (*rpl*33). The four genes not included in fig.2 either show only a trinucleotide sequence for potential base pairing (*atp*B) or contain out-of-frame initiator codons between the potential recognition sites and the respective in frame start codons (*rps*16, *rpo*B and *pet*D). In the leader sequence AGAGGAAUGUUAUG of *rps*16 mRNA an out-of-frame AUG immediately follows the potential recognition sequence, while the in-frame AUG is 5 nucleotides distal to it. This situation possibly favours the translational start at the in-frame AUG (reminiscent of the situation in several bacterial mRNAs [8]) and, with this reservation, the *rps*16 leader could have been included in fig.2. In view of other potential initiator codons for the *rpo*B reading frame the situation remains open for final identification of a recognition sequence. The vast majority of the mRNAs encoded in the tobacco plastom, however, do contain potential recognition sites, which gives support to the conclusion [3] that a mechanism similar to the one proposed [4], and experimentally verified [5–7] for bacterial translation, is also functioning in chloroplasts from higher plants.

In contrast to bacterial mRNAs, where ribosome recognition sites are separated by 7 ± 2 nucleotides from their respective initiator codons [8], spacing in chloroplast mRNAs appears less uniform. As is evident from fig.2, spacing varies between a single nucleotide position in the *pet*B gene and 25 nucleotides in the *rpl*2 gene. In the case of the *pet*B mRNA an in-frame GUG codon is located 7 nucleotides from the potential recognition sequence. It appears, therefore, likely that this GUG instead of the nearer AUG is the functional initiator codon. A frequency distribution of the individual positions potentially involved in base pairing with 16 S rRNA is presented in the lower part of fig.2. It shows a strong peak occurring around position −8, which is in close accordance with the average distance observed in bacterial mRNAs. However, a shoulder around position −15 and an additional peak around position −23

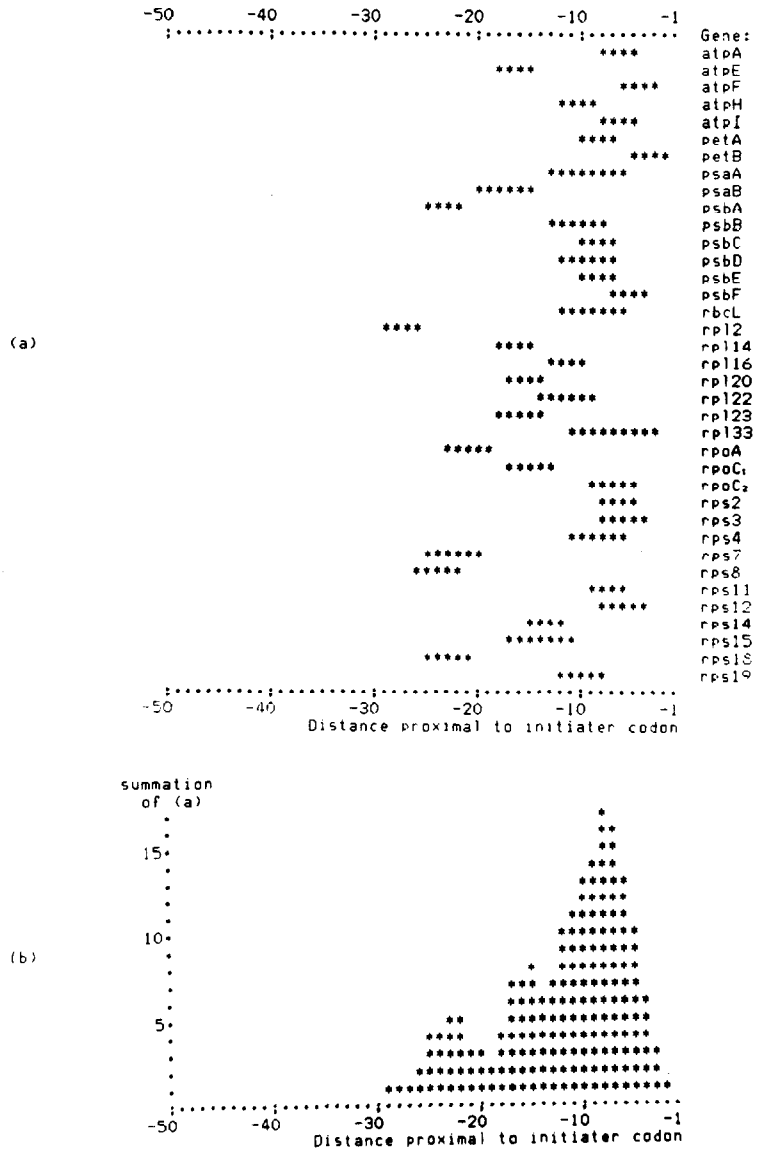

Fig.2. (a) Position and length of mRNA leader sequences complementary to the 3′-end of 16 S rRNA from 37 mRNAs encoded in the tobacco plastome. (b) Frequency distribution of the positions listed in (a).

are clearly visible. From this it appears likely that chloroplast ribosomes can accommodate larger distances between the recognition sequences and their respective initiator codons during translational initiation. Taking into account a wider spacing, recognition sequences can also be identified for mRNAs such as psbA mRNAs, from different species which were thought to lack or to have poor recognition sites [11,12].

It should, however, be emphasized that almost all the initiator codons used for this screening were deduced from DNA sequences alone. Therefore the possibility should be kept in mind that different initiator codons – either further downstream in frame or newly generated by splicing [15] may function as recognition sequences not identified in this screening. Whether the recognition sites further removed from the initiator codons are

truly functional remains to be verified experimentally by site-specific mutation and in vitro translational tests using a chloroplast-derived system.

Acknowledgements: We thank Dr M. Sugiura for providing us with a computer-readable form of the tobacco chloroplast DNA sequence. Critical reading of the manuscript by Dr G. Igloi is gratefully acknowledged. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 206) and by the Fonds der Chemischen Industrie, FRG.

REFERENCES

[1] Schwarz, Z. and Kössel, H. (1980) Nature 283, 739–742.
[2] Dams, E., Hendriks, L., Van de Peer, Y., Neefs, J.-M., Smits, G., Vandenbempt, I. and De Wachter, R. (1988) Nucleic Acids Res. 16, r87–r173.
[3] Schwarz, Z. and Kössel, H. (1979) Nature 279, 520–522.
[4] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[5] Steitz, J.A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. USA 72, 4734–4738.
[6] Jacob, W.F., Santer, M. and Dahlberg, A.E. (1987) Proc. Natl. Acad. Sci. USA 84, 4757–4761.
[7] Hui, A. and De Boer, H.A. (1987) Proc. Natl. Acad. Sci. USA 84, 4762–4766.
[8] Kozak, M. (1983) Microbiol. Rev. 47, 1–45.
[9] Krebbers, E.T., Larrinua, I.M., McIntosh, L. and Bogorad, L. (1982) Nucleic Acids Res. 10, 4985–5002.
[10] Zurawski, G., Perrot, B., Bottomley, W. and Whitfeld, P.R. (1981) Nucleic Acids Res. 9, 3251–3270.
[11] Sugita, M. and Sugiura, M. (1984) Mol. Gen. Genet. 195, 308–313.
[12] Zurawski, G., Bohnert, H.J., Whitfeld, P.R. and Bottomly, W. (1982) Proc. Natl. Acad. Sci. USA 79, 7699–7703.
[13] Ruf, M. and Kössel, H. (1988) Nucleic Acids Res. 16, 5741–5754.
[14] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
[15] Tanaka, M., Obokata, J., Chunwongse, J., Shinozaki, K. and Sugiura, M. (1987) Mol. Gen. Genet. 209, 427–431.
[16] Hudson, G.S., Holton, T.A., Whitfeld, P.R. and Bottomley, W. (1988) J. Mol. Biol. 200, 639–654.
[17] Sijben-Müller, G., Hallick, R.B., Alt, J., Westhoff, P. and Herrmann, R.G. (1986) Nucleic Acids Res. 14, 1029–1044.